applied sciences

MDPI

Article

Real-Time Analysis of Facial Expressions for Mood Estimation

Juan Sebastián Filippini †, Javier Varona *,† and Cristina Manresa-Yee †

Unitat de Gràfics i Visió per Ordinador i IA, Department of Mathematics and Computer Science, University of the Balearic Islands, 07122 Palma, Spain; jseba@mpl.cat (J.S.F.); cristina.manresa@uib.es (C.M.-Y.)

* Correspondence: xavi.varona@uib.es

† These authors contributed equally to this work.

Abstract: This paper proposes a model-based method for real-time automatic mood estimation in video sequences. The approach is customized by learning the person's specific facial parameters, which are transformed into facial Action Units (AUs). A model mapping for mood representation is used to describe moods in terms of the PAD space: Pleasure, Arousal, and Dominance. From the intersection of these dimensions, eight octants represent fundamental mood categories. In the experimental evaluation, a stimulus video randomly selected from a set prepared to elicit different moods was played to participants, while the participant's facial expressions were recorded. From the experiment, Dominance is the dimension least impacted by facial expression, and this dimension could be eliminated from mood categorization. Then, four categories corresponding to the quadrants of the Pleasure–Arousal (PA) plane, "Exalted", "Calm", "Anxious" and "Bored", were defined, with two more categories for the "Positive" and "Negative" signs of the Pleasure (P) dimension. Results showed a 73% of coincidence in the PA categorization and a 94% in the P dimension, demonstrating that facial expressions can be used to estimate moods, within these defined categories, and provide cues for assessing users' subjective states in real-world applications.

Keywords: affective analysis; mood; facial expressions; computer vision; visual tracking

Citation: Filippini, J.S.; Varona, J.; Manresa-Yee, C. Real-Time Analysis of Facial Expressions for Mood Estimation. *Appl. Sci.* **2024**, *14*, 6173. https://doi.org/10.3390/app14146173

Academic Editors: Yuan Zong, Xin Liu and Jingang Shi

Received: 28 May 2024
Revised: 1 July 2024
Accepted: 5 July 2024
Published: 16 July 2024

1. Introduction

Nowadays, the recognition of facial expressions is a very dynamic field, as it covers a wide range of applications in psychology [1] and advertising or marketing [2], among others [3]. This recognition is frequently performed according to the Facial Action Coding System (FACS) [4,5]. FACS allows analyzing human facial expressions through facial coding, and it can be used to classify any anatomical facial expression by studying the movements of the muscles associated with a facial expression. These movements are divided into what is commonly referred to as Action Units (AUs), which are the fundamental actions of muscles or individual muscle groups (e.g., AU6 refers to raising cheeks).

Additionally, the terms mood and emotion are normally confused in colloquial language and in their formal definitions. However, there is a general consensus nowadays that establishes the main differences between both terms [6]. Mainly, mood lasts longer than emotions, and it is not outwardly expressed in facial expressions in a direct manner. Instead, mood is related to emotions insofar as a person who is in a certain mood tends to experience certain emotions. In other words, by means of noticeable effects produced by emotions, facial expressions, or gestures, it is possible to recognize a person's mood. This work focuses on mood, specifically on its recognition from facial expressions.

In early works, the recognition of affect has been achieved mainly as a post-processing stage where images are segmented, analyzed, and finally the identified emotional state is given. Previous works have successfully achieved the automatic detection of the six basic, or universal, emotions in controlled environments [7–9]. Nevertheless, when it comes to the continuous and spontaneous recognition of affect, there are still several challenges given that the input (visual appearance, audio profile, or spontaneous behavior)

differs from the controlled behaviors [10]. In this line, McDuff et al. [11] provided a reliable valence classification based on spontaneous head and facial AUs, categorizing them into positive, negative, and neutral valence. They observed users' spontaneous behaviors while watching various movie clips and compared different classifiers. Similarly, Nicolaou et al. [12] proposed a novel Ouput-Associative Relevance Vector Machine (OA-RVM) regression framework, resulting in a more accurate and robust model that could be successfully used in dimensional (arousal and valence) and continuous emotion prediction from facial expressions, shoulder movements and audio cues. The preceding studies primarily concentrate on emotion recognition or the estimation of dimensional mood in terms of valence and arousal, rather than on a comprehensive categorization of mood.

In the last few years, researchers have been putting their efforts in obtaining more accurate recognition of continuous and spontaneous behaviors [13]. Considering the recent trend towards the continuous and multimodal prediction of spontaneous affective displays in the wild, deep learning is generally well suited to address the challenges faced by such systems. However, in comparison to related fields like object recognition, the impact of deep learning on affect recognition has not yet been fully felt [14,15]. One of the most significant drawbacks of deep learning is that it requires more processing time due to the large amount of data it must deal with. This could be overcome by employing various parallel hardware architecture platforms, such as GPU and FPGA [16]. In addition, there is a higher difficulty for the inherent ambiguity of annotations in affective databases [17].

In contrast, this study introduces a mood data set annotated directly by users through a self-assessment questionnaire. The data set presents video examples of natural reactions to audio-visual stimuli which are similar to the ones used to conduct market research studies with focus groups. However, our data set (composed of 69 sequences) is small in order to use deep learning techniques, but it is suitable for traditional computer vision ones. Therefore, we use a traditional approach based on facial feature tracking and the learning of the particular user expressions. An advantage of our proposal versus deep learning methods is that all of the process is explainable and interpretable, understanding how decisions are made. In addition, our method has been implemented on simpler and less expensive hardware, such as smartphones, which makes its implementation more accessible and affordable than deep learning methods. The main contribution of this study is the development of a real-time, vision-based method for mood estimation using a categorization in the Pleasure–Arousal–Dominance (PAD) space. This method offers valuable insights for evaluating users' subjective emotional states in practical, real-world settings.

In the following sections we present our method of mood estimation and its performance evaluation. Section 2 describes the methodology for the mapping of AUs into the PAD space, while Sections 3 and 4 explain the vision-based estimation of mood and the experiments conducted for capturing the user's affective states produced by different video stimuli. Finally, the last section discusses and concludes this work.

2. Affective States: Emotions and Moods

To the present date, the most studied affective elements are emotions. In fact, the definition of the term affective computing, coined by Rosalind Picard, refers to it as the field that "relates to, arises from, or deliberately influences emotions" [18]. In everyday life people refer to emotions as discrete categories, which have been defined by different psychologists [19–21]. The most popular categorization is the one of basic or universal emotions: joy, sadness, anger, disgust, fear, and surprise. Nonetheless, discrete lists of emotions fail to describe the range of emotions that occur in natural communication settings. An alternative is a dimensional description [22–24], where an affective state is characterized in terms of a number of dimensions rather than in terms of a small number of discrete categories. These dimensions include evaluation, activation, control, power, among others. The idea originated from the observation that some emotions share characteristics that can be seen

as different degrees of two or more dimensions. Therefore they do not need to be labelled and categorized, constraining their study and measurement.

There are also other affective traits that have been explored in affective computing, among which we find mood. According to Sedikides [25], mood states are defined as "frequent, relatively long and pervasive, but typically milder in intensity than emotions". Others like Neumann et al. [26] suggest that "pre-existing mood increases the intensity of affectively congruent emotions while dampening the intensity of incongruent emotions independent of attributional knowledge". Thus, it is noted that the main difference between emotions and moods is the temporal nature of the latter.

As it can be seen, some of the works that dealt with multimodal and continuous recognition of affect used the dimensional Pleasure–Arousal (or Valence–Arousal) space for emotion representation. Instead, other researchers used the extended Pleasure–Arousal–Dominance space (PAD), proposed by Mehrabian [27] for emotion recognition. Among these works we find the one of Arifin and Cheung [28], who introduced a method for affect-based video segmentation taking as cues the visual and background music information of the videos.

The PAD model is a framework that allows the definition and measurement of different emotional states, emotional traits, and personality traits in terms of three nearly orthogonal dimensions: Pleasure (P), Arousal (A), and Dominance (D). The dimension Pleasure–displeasure distinguishes positive affective states from negative ones. Arousal is defined in terms of a combination of mental alertness and physical activity. Dominance–submissiveness is defined in terms of control versus lack of control over events, one's surroundings, or other people. Thus, from the intersection of the Pleasure, Arousal, and Dominance axes, eight octants can be derived, which represent mood categories (Table 1).

Table 1. Octants in the PAD space: Moods.

Exuberant (+P +A +D)	Bored (−P −A −D)
Docile (+P −A −D)	Hostile (−P +A +D)
Dependent (+P +A −D)	Disdainful (−P −A +D)
Relaxed (+P −A +D)	Anxious (−P +A −D)

The PAD model can be used for defining moods and it allows their interrelation with the facial coding in FACS. In other words, PAD can describe a mood in terms of AUs. Particularly, Arellano et al. [29] defined the correspondence between AUs and the PAD space octants by means of the PAD model. The main objective of this correspondence was the description of each of the eight moods in AU terms. The Facial Expression Repertoire is known for this description [30]. From this mapping a set of rules were obtained, which computed the activation areas and intensities of each AU in PAD. As a result, they obtained a general mapping of AUs into regions of the PAD space.

3. Vision-Based Mood Estimation

This Section describes the vision-based mood estimation method. Figure 1 shows the flowchart of the processing steps of the proposed method:

1. Detect the characteristic facial points of the subject in each image frame to identify the AUs corresponding to the movement of the facial points with respect to the resting patterns.
2. Define a resting pattern corresponding to the distances between the characteristic facial points of the subject.
3. Obtain, for each image of the sequence, the activation probability distribution of the AUs and determine the similarity between the probability distributions for each possible mood to estimate mood as the maximum of the computed similarities.

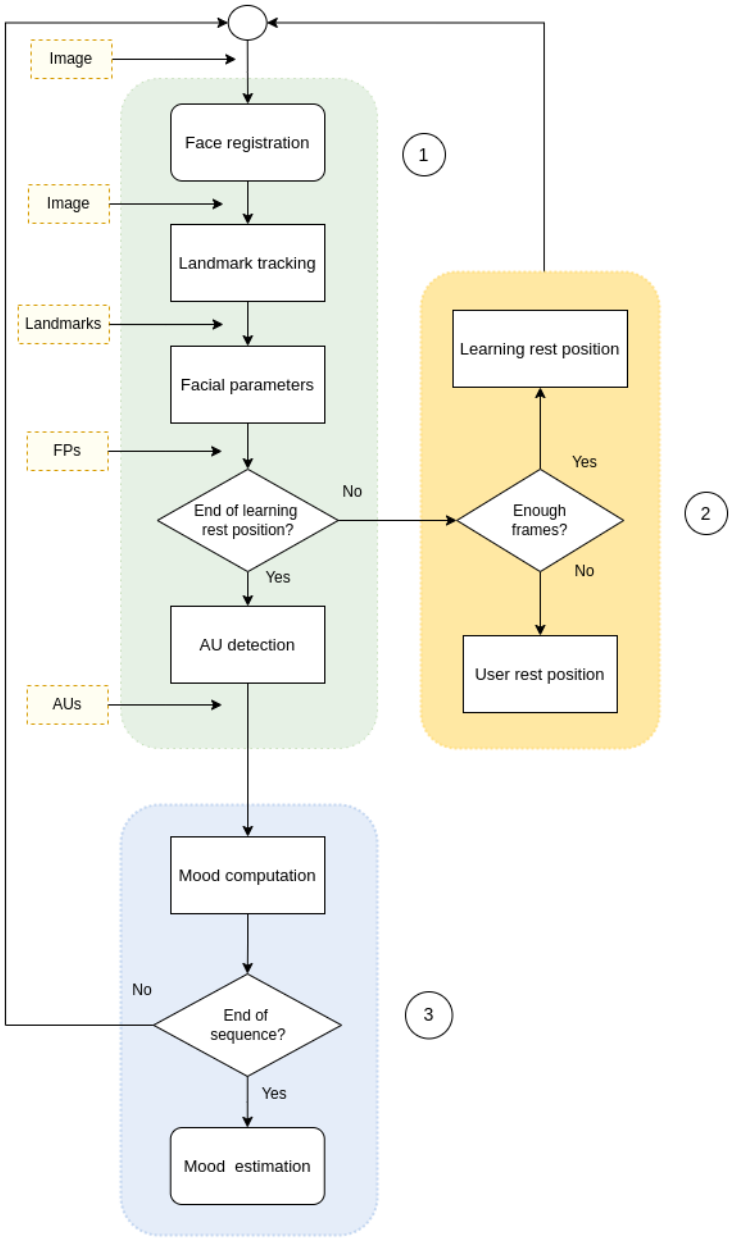


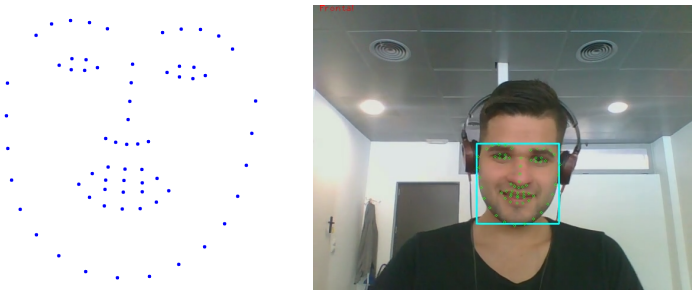
Figure 1. Flowchart with the steps of processing.

3.1. Identification of the Activation of Action Units

As described in the previous section, AUs can be used for the interpretation of facial signals of mood. The former results led us to consider the use of this mapping in the estimation of facial expressions of mood. The motivation behind this is that it would be useful to know not only the emotions that the person is experiencing in an instant of time, but also their mood during certain interval of time. This could open a wide range of new possibilities for human–computer interaction (HCI) applications that deal with the affective state of the user.

Saragih et al. developed a generic person-independent facial tracker that exhibits a high degree of robustness towards variations frequently encountered in real images [31]. We use this tracker to capture the motion of facial points, see Figure 2. This method uses a parametric shape model, which captures the statistical variation of a set of landmarks on the facial image. This model is built using Principal Component Analysis of training images. Next, for each landmark, a function returns a response map indicating the probability that the landmark is located in a certain position based on the result of a previous iteration

or an initialization value. To estimate the final position of the landmarks, an adjustment is performed that jointly searches for the maxima in the response maps, maintaining the constraints given by the shape model. This centrally characterizes the problem, which is why the authors define it as a constrained local model (CLM). The advantage over other proposals of facial features tracking is that this tracker robustly captures the motion of 66 facial points.



(**a**) (**b**)

Figure 2. (**a**) Facial landmarks. (**b**) Face tracker.

Similarly to [7], from the tracked points we develop a set of distance-based rules to recognize the AUs. First, we define a set of facial parameters based on distances between tracked points corresponding to facial features, see Table 2.

Table 2. Facial parameters extracted from the tracked facial points.

Facial Parameter	Description
FP1	Inner eyes distance (SCALE)
FP2	Mean eyebrow–eyeline distance (right)
FP3	Inner eyebrow–eyeline distance (right)
FP4	Mean eyebrow–eyeline distance (left)
FP5	Inner eyebrow–eyeline distance (left)
FP6	Open eye distance (right)
FP7	Open eye distance (left)
FP8	Horitzontal mouth width
FP9	Horitzontal mouth height
FP10	Vertical distance between mouth and nouse
FP11	Vertical distance between chin and nouse
FP12	Lips distance

In this case, our strength is that, in comparison with the classical used data sets such as the CK+ and the MMI, we work with long sequences (more than 90 s), typical of a real study of spontaneous expressions. Therefore, the facial parameters for the AUs recognition rules are automatically learned from the tracked points positions in the first seconds of each sequence, when the user rests its face in neutral position. That is, let m be the number of frames used for learning (typically 25 frames, i.e., 1 s) and $FPi = \{FPi^{(1)}, \ldots, FPi^{(m)}\}$ the samples for the i-th facial parameter. We compute the maximum likelihood values for the mean, μ_i, and the variance, σ_i^2, of a probabilistic gaussian density function, $\mathbf{p}(FPi) = G(\mu_i, \sigma_i^2)$, which models the rest values of the facial parameters, by means of Equations (1) and (2).

$$\mu_i = \frac{1}{m} \cdot \sum_{j=1}^{m} FPi^{(j)}. \tag{1}$$

$$\sigma_i^2 = \frac{1}{m} \cdot \sum_{j=1}^{m} (FPi^{(j)} - \mu_i)^2. \tag{2}$$

This model is used to detect that a significant change was produced in the facial parameter if $\mathbf{p}(FPi) < 0.05$, where $\mathbf{p}(FP_i)$ is computed from Equation (3).

$$\mathbf{p}(FPi) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \exp - \frac{(FPi - \mu_i)^2}{2\sigma_i^2}. \quad (3)$$

Note that the parameter FP1 (Inner eyes distance) is used to scale the distances in order to avoid that user motions alter the learned parameters. Once scaled, we use the rules defined in Table 3 to recognize the AUs.

Table 3. Rules for AUs recognition from the facial parameters.

Action Unit	Probabilistic Rule
AU1	$\mathbf{p}(FP3) < 0.05$ & $\mathbf{p}(FP5) < 0.05$ & $FP5 > \mu_5$
AU2	$\mathbf{p}(FP2) < 0.05$ & $\mathbf{p}(FP4) < 0.05$ & $FP4 > \mu_4$
AU4	$\mathbf{p}(FP2) < 0.05$ & $\mathbf{p}(FP4) < 0.05$ & $FP4 < \mu_4$
AU5	$\mathbf{p}(FP6) < 0.05$ & $\mathbf{p}(FP7) < 0.05$ & $FP7 > \mu_7$
AU6	$\mathbf{p}(FP8) < 0.05$ & $\mathbf{p}(FP6) < 0.05$ & $\mathbf{p}(FP7) < 0.05$ & $FP7 < \mu_7$
AU10	$\mathbf{p}(FP10) < 0.05$ & $FP10 > \mu_{10}$
AU12	$\mathbf{p}(FP8) < 0.05$ & $FP8 > \mu_8$
AU15	$\mathbf{p}(FP12) < 0.05$ & $FP12 > \mu_{12}$
AU25	$\mathbf{p}(FP9) < 0.05$ & $FP9 > \mu_9$
AU26	$\mathbf{p}(FP11) < 0.05$ & $FP11 > \mu_{11}$
AU43	$\mathbf{p}(FP8) < 0.05$ & $\mathbf{p}(FP6) < 0.05$ & $\mathbf{p}(FP7) < 0.05$ & $FP7 < \mu_7$

To evaluate the proposed AUs detection in video sequences, we used the MMI database. MMI is a continuously developing online searchable database dedicated to the recognition of AUs and basic emotions from video sequences [32]. MMI contains video sequences annotated according to the FACS system. Some example sequences from the data set are shown in Figure 3.

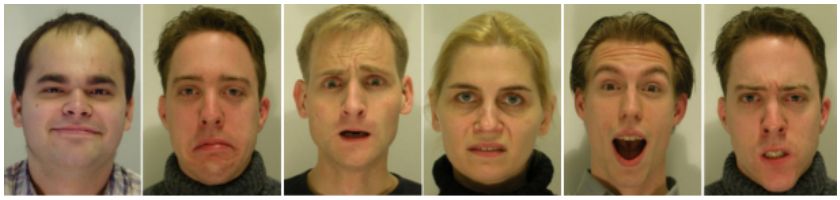


Figure 3. MMI data set examples. From left to right: happiness, sadness, fear, disgust, surprise, and anger.

Unlike other widely used databases, such as the one by Cohn–Kanade ([33,34]) in its extended version, known as CK+, the sequences are complete in terms of the temporal development of facial expression and are presented as video streams at 25 FPS. CK+ sequences, for example, are recorded at 1 FPS and show the development of facial expression from the neutral position to its apex, omitting the subsequent deactivation phase. MMI sequences contain two types of annotation: the list of AUs active at each point during the sequence, and the OAO Annotation (onset–apex–offset). The OAO annotation indicates the list of active AUs at some point during the sequence. The database contains 329 sequences of this type of annotation, from which 252 contain annotations with the AUs studied in this work. However, as we have explained before, our method requires an initial calibration time of 1 s in which all AUs are assumed to be in neutral phase. To ensure compliance with this condition, only sequences with OAO annotation were used, since it is necessary to know the initial time in neutral phase of the sequences. The sequences in which this condition is met were selected, obtaining a set of 65 sequences (6557 frames) on which the validation was performed using the Specificity and Accuracy metrics. The results are

presented in Table 4, where the accuracy and specificity values for the detection of each AU are shown.

Table 4. Action Units detection evaluation results.

AU	TP	FP	TN	FN	Spec	Acc
AU1	31	669	5338	4	**88.42**	**88.42**
AU2	27	871	5171	3	**85.58**	**85.61**
AU4	168	306	5525	73	**94.75**	**93.76**
AU5	58	238	5761	15	**96.03**	**95.83**
AU6	252	732	4981	107	**87.19**	**86.18**
AU10	243	1306	4518	5	**77.58**	**78.41**
AU12	434	1128	4435	75	**79.72**	**80.19**
AU15	7	676	5367	22	**88.81**	**88.50**
AU25	775	1139	3792	366	**76.90**	**75.21**
AU26	216	1251	4433	172	**77.99**	**76.56**
AU43	272	344	5361	95	**93.97**	**92.77**
Total	2483	8690	54,682	937	**86.29**	**85.59**

Table 4 shows that high specificity values are obtained for the entire set of AUs, even though the test set is highly unbalanced in favor of the negative samples. This is important for the mood estimation as false positives have a high impact on its detection. The accuracy is also maintained at acceptable values for the entire sample.

3.2. Mood Estimation

In order to build a robust mood estimation, we define each mood pattern as a standard probability distribution associated with the activation of the AUs associated with each mood (see Table 5).

Table 5. AUs describing facial expressions of mood (from Arellano et al. [29]).

Mood	Action Units (AUs)
Exuberant	AU6, AU5, AU12, AU25, AU26
Bored	AU1, AU2, AU4, AU15, AU43
Docile	AU1, AU2, AU12, AU43
Hostile	AU4, AU10, AU5, AU15, AU25, AU26
Anxious	AU1, AU2, AU4, AU5, AU15, AU25, AU26
Relaxed	AU6, AU12, AU43
Dependent	AU1, AU2, AU5, AU12, AU25, AU26
Disdainful	AU4, AU15, AU43

Let $p_{ij} \in [0, 1]$ be the contribution of each AU j to the mood i, where if an AU is highly determinant, it is assigned the value 1, whereas if it is not important for a certain mood, it is assigned the value 0. Then, a pattern, $\mathbf{p}_i$, is constructed for each mood with the following values as the AU contributions for the i-th mood:

$$\mathbf{p}_i = C_p \cdot (p_{i1}, p_{i2}, ..., p_{in}), \tag{4}$$

where C_p is a normalization constant for imposing the condition that $\sum_j p_{ij} = 1$, and n is the number of AUs.

From a new video sequence of W images, which are consecutive in the sequence, the activated AUs can be determined by means of the previously described method (see Section 3.1). By repeating that comparison with all the images, it is possible to determine if an AU has been activated in one or in several frames. In other words, an occurrence or

relevance value, can be obtained for each AU. Each of those occurrence values for each AU_j is referred to as q_j, which is calculated by means of the expression of Equation (5).

$$q_j = \frac{C_q}{W} \cdot \sum_{k=1}^{W} s_{kj}. \tag{5}$$

Equation (5) is a normalized mean of the AU_j ocurrence in the complete video. That is, s_{kj} represents the activation or non-activation of the corresponding AU_j in the image corresponding to frame k. If the AU_j has been activated, s_{kj} is assigned the value $s_{kj} = 1$, whereas if it has not been activated, $s_{kj} = 0$. Finally, C_q is a normalization constant for imposing the condition that $\sum_{j=1}^{n} q_j = 1$.

For determining the similarity between the estimated AU_j distribution, q_j, and each mood pattern, $\mathbf{p}_i$, the Bhattacharyya coefficient [35], D_i, is used for each mood i according to Equation (6).

$$D_i = \sum_{j=1}^{n} (p_{ij} \cdot q_j)^{1/2}. \tag{6}$$

Finally, the estimated mood is the maximum of the computed similarities, $M = \max_i D_i$.

A reliable and robust mood recognition method is thereby achieved, where image analysis is performed in sequences captured by the camera, which allow for evaluating dynamically the contribution of the AUs to the mood of the subject. In addition, as previously mentioned (see Section 2), the PAD model is defined in terms of three orthogonal dimensions: Pleasure (P), Arousal (A), and Dominance (D). Thus, from the intersection of the axes of Pleasure, Arousal, and Dominance, eight octants can be derived that represent the fundamental categories of mood. The mood estimation fits one of these eight fundamental categories and requires knowing the sign in each of the dimensions of the PAD space, which, depending on the expressiveness of the subject, may not be possible. As defined in Mehrabian's original work, Dominance is more related to an internal feeling of control or submission, which may not have obvious or consistent facial manifestations. In the context of image sequence analysis for emotion recognition, most previous work focuses on identifying facial expressions that correspond to basic emotion categorizations or the Pleasure (Valence) and Arousal dimensions [14]. In contrast, Dominance may require additional contextual or postural information to be accurately inferred, so a second level of categorization was considered, eliminating this dimension. This second level is made up of four categories corresponding to the quadrants of the P+A plane: "Exaltation", "Calm", "Anxiety" and "Bored". Following the same criteria, a third level corresponding to the P dimension was defined. This level is made up of two categories: "Positive" and "Negative". Mood states for which it is not possible to determine the P value are considered "Neutral". Figure 4 shows a selection of frames from a sequence with different expression activation, the identified AUs and the estimated mood for each level. The maximum values of each level are reported as the final mood estimation.

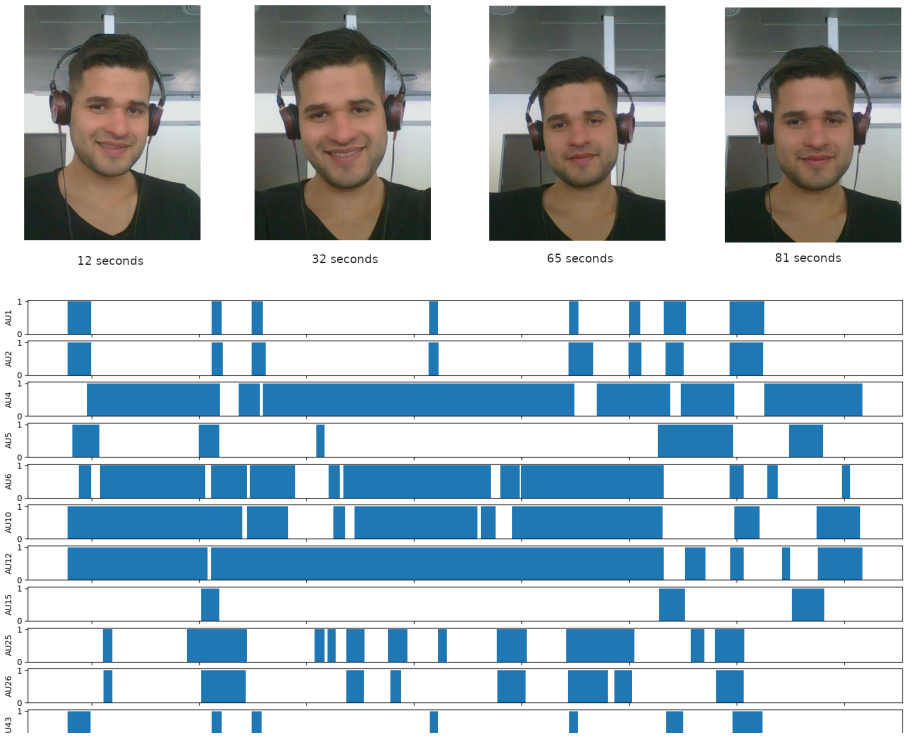


Figure 4. Example of the sequence processing, including the AU identification for each frame. The results of the different levels of mood estimation are: L1—Exhuberant, L2—Elated, L3—Positive.

4. Performance Evaluation

Most existing affective data sets consist of posed static images or acted expression recordings. Therefore, to evaluate the mood estimation method, we built a data set that contains video examples of natural reactions to audio-visual stimuli and participants self-reported their affective state.

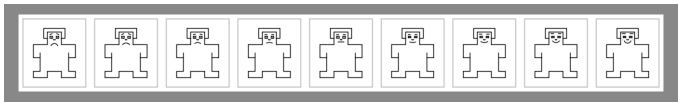
4.1. Material: Data Set for Evaluation

To build the data set, 60 individuals (30 females) participated without being instructed to pose specific expressions or emotions. Data collection took place in natural environments, such as the participants' workplaces or homes, rather than in a laboratory setting. Consequently, the resulting video sequences exhibit diversity in background and visual distractors like non-frontal faces and the presence of facial artifacts (e.g., glasses or beards), see Figure 5.

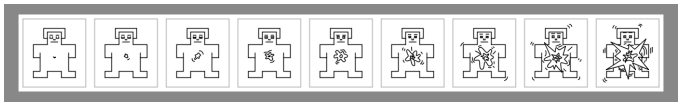
Figure 5. Frames of the different participants of the mood data set including examples of visual distractors.

Participation was entirely voluntary, and participants were informed that the data set would be used exclusively for research purposes. Additionally, they were assured that all collected information would be kept private, anonymous, and confidential.

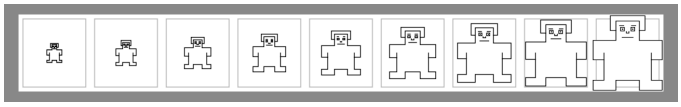
To evoke different affective states, a set of stimuli video, lasting from one to four minutes each video, was prepared to elicit one of the eight mood categories. As a reference, the International Affective Picture System (IAPS) was used [36]. IAPS is a static image data set intended to elicit certain emotions. The IAPS data set contains 1182 images that were rated during a period of ten years by males, females and children using the Self-Assessment Manikin (SAM) questionnaire, a non-verbal pictorial assessment technique that directly measures the PAD dimensions associated with a person's affective reaction. Figure 6 shows icons used for each PAD dimension in the SAM questionnaire.



(**a**) Pleasure scale. Left icons represent less pleasure.



(**b**) Arousal scale. Left icons represent less arousal.



(**c**) Dominance scale. Left icons represent less situation control.

Figure 6. Nine points SAM scales.

The procedure to record each video followed the below steps:

1. The experiment was explained to the participant, mainly the questionnaire to be completed in the last step. The participant signed the informed consent and demographic data were gathered.
2. A randomly selected video from the set prepared to elicit one of the eight mood categories was presented to the participant. A maximum of two videos was shown to the participants.
3. While the stimulus video was playing, the participant's face was recorded. The only recommendation given to the participant was to remain moderately centered in front of the camera.
4. At the end, participants completed the SAM questionnaire to obtain the mood values in the PAD space. Because PAD levels were difficult to understand for some participants, a post-processing stage was carried out, and sequences with strong discrepancy regarding the stimuli were discarded.

Finally, we obtained a data set with 69 video sequences and we enriched the database metadata with the subjects' age and gender reported by themselves.

4.2. Procedure

To evaluate the performance of the mood estimator, we compared the participants' self-rated affective states for each video with the estimator's predictions. The analysis was conducted across three levels of affective dimensions:

First level. Positive and Negative.

Second level. Exalted, Calm, Anxious, and Bored.

Third level. Elated, Dependent, Relaxed, Docile, Hostile, Anxious, Bored, and Disdainful.

We counted the number of matches between the self-ratings and the estimator's outcomes for each dimension level.

4.3. Results

The results of the comparison are presented in Table 6. At the first level (Positive and Negative), there was a 94% match between the estimated mood and the self-rated values from the SAM questionnaire. At the second level (Exalted, Calm, Anxious, and Bored), a 73% match was observed. Finally, at the third level (Elated, Dependent, Relaxed, Docile, Hostile, Anxious, Bored, and Disdainful), the match between the annotations and the estimated mood was 64%. It is noticeable that, in general, regarding the cases for which there is no match, we can mainly observe tracking errors caused by the presence of facial artifacts such as beards or glasses, and a lack of facial expressiveness.

Table 6. Mood estimation evaluation results.

Mood (SAM)	Dimensions	Coincidences	Sequences
Positive	P+	32	33
Negative	P−	33	36
Level 1		**65**	**69**
Exalted	P+, A+	14	21
Calm	P+, A−	11	12
Anxious	P−, A+	12	16
Bored	P−, A−	14	20
Level 2		**51**	**69**
Elated	P+, A+, D+	11	15
Dependent	P+, A+, D−	3	6
Relaxed	P+, A−, D+	9	9
Docile	P+, A−, D-	1	3
Hostile	P−, A+, D+	3	5
Anxious	P−, A+, D−	8	11
Disdainful	P−, A−, D+	2	7
Bored	P−, A−, D−	7	13
Level 3		**44**	**69**

5. Limitations

The current method is designed to function under specific acquisition conditions to operate in real-time applications effectively. The user's face needs to be frontal and fully visible within the image, although slight pose variations are tolerated. However, significant deviations may lead to a loss of tracking. While the system can accommodate occasional occlusions, these, along with tracking recovery time, reduce the analysis time and, therefore, can impact the accuracy of mood estimation. Addressing these limitations is a focus for future improvements.

6. Conclusions

This paper presented a novel approach for the real-time estimation of mood from facial expressions. The main novelty regards the mapping of AUs into the PAD space, obtaining consistent sets of AUs that describe different expressions for each mood. The other contribution is the real-time recognition of mood in spontaneous facial expressions in long video sequences.

We also present a mood data set annotated by the users themselves by means of a self-assessment questionnaire. The data set contains video examples of natural reactions to audio-visual stimuli similar to the ones used to conduct research studies with focus groups. The results obtained in the conducted experiments show that Dominance is the

dimension least impacted by facial expression, and this dimension could be eliminated from mood categorization. Then, four categories corresponding to the quadrants of the Pleasure–Arousal plane, "Exalted", "Calm", "Anxious" and "Bored" were maintained, with two categories of "Positive" and "Negative" signs for the Pleasure (P) dimension. The 73% of coincidence in the Pleasure–Arousal categorization and the 94% of coincidence in the Pleasure dimension with the questionnaires filled by the participants show that mood can be indeed recognized automatically and perceived subjectively in the particular conditions of research studies.

This research represents a step forward in the recognition of affect in facial expressions because it is not limited to the common use of expressions of emotions, but opens a new door for the recognition of different affective traits based on the PAD model. In addition, the presented method, which works in real time and is based on traditional computer vision techniques, could be implemented on less expensive hardware, such as a smartphone, making its implementation more affordable. The source code for the Android OS version is available at the following GitHub repository: https://github.com/jsebaf/facetomood, accessed on 1 July 2024.

In the future, we intend to broaden the set of mapped AUs for a finer detection, as well as the mood estimation in other environments where spontaneous behavior is possible.The ability of recent generative deep learning models to generate synthetic images of faces could address the problem in another way. The possibility of controlling the facial latent space to model facial expressions opens the hypothesis of exploring better representations of facial expressions to facilitate the estimation of affective states since, currently, there is only a certain guarantee of results in the recognition of microexpressions.

Author Contributions: Conceptualization, J.V.; methodology, J.S.F., J.V. and C.M.-Y.; software, J.S.F.; validation, J.S.F., J.V. and C.M.-Y.; formal analysis, J.S.F. and J.V.; investigation, J.S.F., J.V. and C.M.-Y.; data curation, J.S.F.; writing—original draft preparation, J.S.F., J.V. and C.M.-Y.; writing—review and editing, J.S.F., J.V. and C.M.-Y.; visualization, J.S.F.; supervision, J.V. and C.M.-Y.; project administration, J.S.F. and J.V.; funding acquisition, C.M.-Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work is part of the Project PID2022-136779OB-C32 (PLEISAR) funded by MCIN/AEI/10.13039/501100011033/ and FEDER: "A way to make Europe".

Institutional Review Board Statement: Since the presented study is non-interventional, all participants gave written informed consent and are aware that their participation is voluntary and that their images will be used for research purposes, it does not require the approval of the Ethics Committee.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are registered in the University of the Balearic Islands repository and are available on request from the corresponding author.

Acknowledgments: The authors acknowledge the University of the Balearic Islands, and the Department of Mathematics and Computer Science for their support.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Grabowski, K.; Rynkiewicz, A.; Lassalle, A.; Baron-Cohen, S.; Schuller, B.; Cummins, N.; Baird, A.; Podgórska-Bednarz, J.; Pieniążek, A.; Łucka, I. Emotional expression in psychiatric conditions: New technology for clinicians. *Psychiatry Clin. Neurosci.* **2019**, *73*, 50–62. [CrossRef] [PubMed]
2. Barreto, A.M. Application of facial expression studies on the field of marketing. In *Emotional Expression: The Brain and the Face*; FEELab Science Books: Porto, Portugal, 2017; Volume 9, pp. 163–189.
3. Sariyanidi, E.; Gunes, H.; Cavallaro, A. Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1113–1133. [CrossRef] [PubMed]
4. Ekman, P.; Friesen, W. *Facial Action Coding System: Manual*; Consulting Psychologists Press: Palo Alto, CA, USA, 1978.
5. Friesen, W.V.; Ekman, P. EMFACS-7: Emotional Facial Action Coding System. *Psychol. Comput. Sci.* **1983** .
6. Ekman, P.; Davidson, R. *The Nature of Emotion*; Oxford University Press: Oxford, UK, 1994.

7. Pantic, M.; Patras, I. Detecting facial actions and their temporal segments in nearly frontal-view face image sequences. In Proceedings of the 2005 IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI, USA, 12 October 2005; Volume 4, pp. 3358–3363.
8. Kotsia, I.; Pitas, I. Facial expression recognition in image sequences using geometric deformation features and support vector machines. *Image Process. IEEE Trans.* **2007**, *16*, 172–187. [CrossRef] [PubMed]
9. Valstar, M.; Pantic, M. Fully Automatic Recognition of the Temporal Phases of Facial Actions. *IEEE Trans. Syst. Man Cybern. Part B* **2011**, 42, 28–43. [CrossRef] [PubMed]
10. Zeng, Z.; Pantic, M.; Roisman, G.; Huang, T. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Anal. Mach. Intell. IEEE Trans.* **2009**, *31*, 39–58. [CrossRef] [PubMed]
11. McDuff, D.; Kaliouby, R.E.; Kassam, K.; Picard, R. Affect valence inference from facial action unit spectrograms. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, USA, 13–18 June 2010; pp. 17–24.
12. Nicolaou, M.A.; Gunes, H.; Pantic, M. Output-associative RVM regression for dimensional and continuous emotion prediction. *Image Vis. Comput.* **2012**, *30*, 186–196. [CrossRef]
13. Wang, Y.; Song, W.; Tao, W.; Liotta, A.; Yang, D.; Li, X.; Gao, S.; Sun, Y.; Ge, W.; Zhang, W.; et al. A systematic review on affective computing: Emotion models, databases, and recent advances. *Inf. Fusion* **2022**, *83–84*, 19–52. [CrossRef]
14. Li, S.; Deng, W. Deep Facial Expression Recognition: A Survey. *IEEE Trans. Affect. Comput.* **2022**, *13*, 1195–1215. [CrossRef]
15. Boughanem, H.; Ghazouani, H.; Barhoumi, W. Facial Emotion Recognition in-the-Wild Using Deep Neural Networks: A Comprehensive Review. *SN Comput. Sci.* **2024**, *5*, 96. [CrossRef]
16. AlBdairi, A.J.A.; Xiao, Z.; Alkhayyat, A.; Humaidi, A.J.; Fadhel, M.A.; Taher, B.H.; Alzubaidi, L.; Santamaría, J.; Al-Shamma, O. Face recognition based on deep learning and FPGA for ethnicity identification. *Appl. Sci.* **2022**, *12*, 2605. [CrossRef]
17. Rouast, P.V.; Adam, M.T.P.; Chiong, R. Deep Learning for Human Affect Recognition: Insights and New Developments. *IEEE Trans. Affect. Comput.* **2021**, *12*, 524–543. [CrossRef]
18. Picard, R.W. *Affective Computing*; MIT Press: Cambridge, MA, USA, 1997.
19. Darwin, C. *The Expression of Emotions in Man and Animals*; Murray: London, UK, 1873.
20. Ekman, P. *Emotion in the Human Face*; Cambridge University Press: Cambridge, UK, 1982.
21. Plutchik, R. *Emotions: A Psychoevolutionary Synthesis*; Harper & Row: New York, NY, USA, 1980.
22. Whissell, C. The dictionary of affect in language. In *Emotion: Theory, Research, and Experience, Volume 4: The Measurement of Emotions*; Plutchik, R., Kellerman, H., Eds.; Academic Press: New York, NY, USA, 1989; Chapter 5, pp. 113–131.
23. Russell, J.A. Measures of emotion. In *Emotion: Theory, Research, and Experience. The Measurement of Emotions*; Plutchik, R., Kellerman, H., Eds.; Academic Press: New York, NY, USA, 1989; Volume 4, Chapter 4, pp. 83–111.
24. Cochrane, T. 8 dimensions for the emotions. *Soc. Sci. Inf. Spec. Issue The Lang. Emot. Concept. Cult. Issues* **2009**, *48*, 379–420.
25. Sedikides, C. Changes in the Valence of the Self as a Function of Mood. *Rev. Personal. Soc. Psychol.* **1992**, *14*, 271–311.
26. Neumann, R.; Seibt, B.; Strack, F. The influence of mood on the intensity of emotional responses: Disentangling feeling and knowing. *Cogn. Emot.* **2001**, *15*, 725–747. [CrossRef]
27. Mehrabian, A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Curr. Psychol.* **1996**, *14*, 261–292. [CrossRef]
28. Arifin, S.; Cheung, P.Y.K. A computation method for video segmentation utilizing the pleasure-arousal-dominance emotional information. In Proceedings of the MM '07: Proceedings of the 15th ACM International Conference on Multimedia , Augsburg, Bavaria, Germany, 23–28 September 2007; pp. 68–77. [CrossRef]
29. Arellano, D.; Perales, F.J.; Varona, J. Mood and Its Mapping onto Facial Expressions. In *Proceedings of the Articulated Motion and Deformable Objects*; Perales, F.J., Santos-Victor, J., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 31–40.
30. Biehn, C. Facial Expression Repertoire (FER). 2005. Available online: http://research.animationsinstitut.de/ (accessed on 18 January 2014).
31. Saragih, J.M.; Lucey, S.; Cohn, J.F. Deformable model fitting by regularized landmark mean-shift. *Int. J. Comput. Vis.* **2011**, *91*, 200–215. [CrossRef]
32. Valstar, M.; Pantic, M. Induced Disgust, Happiness and Surprise: An Addition to the Mmi Facial Expression Database. In *Proc. 3rd Intern. Workshop on EMOTION (Satellite of LREC): Corpora for Research on Emotion and Affect*; 2010; p. 65. Available online: https://ibug.doc.ic.ac.uk/research/mmi-database/ (accessed on 4 July 2024).
33. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 94–101.
34. Kanade, T.; Cohn, J.F.; Tian, Y. Comprehensive database for facial expression analysis. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), Grenoble, France, 28–30 March 2000; pp. 46–53.

35. Kailath, T. The divergence and Bhattacharyya distance measures in signal selection. *IEEE Trans. Commun. Technol.* **1967**, *15*, 52–60. [CrossRef]
36. Bradley, M.M.; Lang, P.J. International Affective Picture System. In *Encyclopedia of Personality and Individual Differences*; Zeigler-Hill, V., Shackelford, T., Eds.; Springer: Cham, Switzerland, 2017.